Factor analysis of the spectral and time behavior of long GRBs

Zsolt Bagoly*, Luis Borgonovo†, István Horváth**, Attila Mészáros‡ and Lajos G. Balázs§

*Dept. of Physics of Complex Systems, Eötvös University, H-1117 Budapest, Pázmány P. s. 1/A, Hungary
†Stockholm Observatory, AlbaNova, SE-106 91 Stockholm, Sweden
**Dept. of Physics, Bolyai Military University, H-1581 Budapest, POB 15, Hungary
‡Astronomical Institute of the Charles University, V Holešovičkách 2, CZ-180 00 Prague 8, Czech Republic
§Konkoly Observatory, H-1525 Budapest, POB 67, Hungary

Abstract. A sample of 197 long BATSE GRBs is studied statistically. In the sample 11 variables, describing for any burst the time behavior of the spectra and other quantities, are collected. The application of the factor analysis on this sample shows that five factors describe the sample satisfactorily. Both the pseudo-redshifts coming from the variability and the Amati-relation in its original form are disfavored.

INTRODUCTION

Factor Analysis (FA) and the Principal Component Analysis (PCA) are powerful statistical methods in the data analysis. [2] showed that the 9 variables (T_{50}, T_{90}, P_{64}, P_{256}, P_{1024}, $\mathscr{F}_1$, $\mathscr{F}_2$, $\mathscr{F}_3$ and $\mathscr{F}_4$) of the BATSE GRBs can be satisfactorily represented by 3 hidden statistical variables. [3] studied the statistical properties of 197 long BATSE GRBs, using 10 statistical variables describing the temporal and spectral properties of GRBs. Performing a PCA they concluded that about 70 % of the total variance of the parameters were explained by the first 3 Principal Components (PCs).

FA assumes that the observed variables can be explained as a linear combination of hidden variables as given by:

$$x = \Lambda f + \varepsilon, \qquad (1)$$

where x marks an observed variable of p dimension, Λ is a matrix of $p \times m$ dimension ($m < p$) and f means a hidden variable of m dimension. The components of Λ are called loadings and those of f factor scores; ε is a noise term. Observation yields x while the quantities on right-hand-side of Eq.1 have to be computed by a suitable algorithm: here we use the Maximum Likelihood (ML) method. An interesting property of the Λ matrix of factor loadings is that, after undertaking an orthogonal transformation (rotation) on it, one gets an other possible factor solution. Rotation is often useful to get a solution, which is much easier to interpret: we use the varimax rotation in our calculations.

TABLE 1. Factors coming from FA with five factors *SSload* is the sum of the squares of the loadings; *PropVar* defines the proportion of *SSload* to the sum of variances of the input variables; *CVar* defines the sum of proportional variances.

Variable	Fact. 1	Fact. 2	Fact. 3	Fact. 4	Fact. 5
$\log T_{90}$	0.52	-0.05	0.16	0.06	0.83
$\log \mathscr{T}_{50}$	0.84	-0.04	0.01	0.34	0.37
$\log \tau$	0.88	-0.01	-0.02	0.24	0.14
$\log V$	0.32	0.05	0.18	0.72	-0.06
$\log \mathscr{S}_{\mathscr{F}}$	0.10	0.05	-0.17	0.46	0.06
$\log \tau_{\text{lag}}$	0.24	0.02	-0.49	-0.28	-0.03
$\log \mathscr{R}E_{\text{pk}}$	-0.05	0.16	-0.11	-0.49	-0.07
$\log F_{1\text{s}}$	-0.03	0.99	0.07	-0.08	-0.04
$\log \mathscr{F}$	0.66	0.57	0.39	0.05	0.15
$\log E_{\text{pk}}$	0.25	0.16	0.85	-0.23	0.01
α	-0.05	-0.03	-0.32	-0.07	-0.23
SSload	2.43	1.37	1.32	1.30	0.95
PropVar	0.22	0.12	0.12	0.12	0.09
CVar	0.22	0.34	0.46	0.58	0.67

THE SAMPLE

Here we apply the FA on the same sample of 197 long GRBs investigated by [3]: for each burst we use the following 11 variables: duration time T_{90}, emission time $\mathscr{T}_{50}$, autocorrelation function (ACF) half-width τ, variability V, emission symmetry $\mathscr{S}_{\mathscr{F}}$, cross-correlation function time lag τ_{lag}, the ratio of peak energies $\mathscr{R}E_{\text{pk}}$, peak flux on 1024 ms scale $F_{1\text{s}}$, fluence $\mathscr{F}$, peak energy E_{pk}, and low frequency spectral index α.

It is worth mentioning here that, similarly to [3], we do not consider the fluence on the highest channel ($> 300\,\text{keV}$) separately, although in [2] this variable alone defined a PC (factor). This choice is motivated by two arguments: first, because usually the fluences on the fourth channel are often vanishing or have great errors ("the values are noisy"); second, as it is noted by [3], in a sample given by long-soft GRBs *only*, this quantity is less important.

RESULTS AND DISCUSSION

The *m* number of factors, which satisfactorily reproduce the original correlation matrix, can be constrained ([7]) by the inequality of $m \leq (2p + 1 - \sqrt{8p+1})/2$, which here gives $m \leq 6.782$. Since the number of factors is an integer, $m = 6$ is the maximum value.

The ML method used here gives also a probability of the null hypothesis, i.e., that the correlation matrix of the observed variables and that reproduced by the factor solution are statistically identical. For 4, 5 and 6 factors, we get for the validity of the null hypothesis the probabilities of 0.000384, 0.0872, and 0.0973, respectively. These calculations show that 5 factors are already sufficient. The choice of 5 factors can be supported from the cumulative variances too.

The **first factor** is defined by τ, $\mathscr{T}_{50}$, T_{90}, and $\mathscr{F}$, i.e., the first factor is given mainly by the temporal properties. The measures τ and $\mathscr{T}_{50}$ are preferred over T_{90}.

The **second factor** is given mainly by F_{1s} and $\mathscr{F}$. Hence, the second factor is related to the observed strength of the burst. The loadings of τ_{lag} and V are negligible, and hence there is no direct support for the luminosity estimators based on these two variables ([11, 12, 10, 6]).

The first two factors are in accordance with [2] claiming that - among fluence, peak flux and duration - two principal components or factors exist.

The **third factor** is mainly driven by E_{pk}. It is interesting that this peak energy (break-energy) in the spectra is appearing so dominantly as a significant variable in the third factor. It emphasizes that the spectrum itself is an important quantity (a trivially expectable result), and in the spectrum E_{pk} itself is a significant descriptor (this is not a triviality). The loadings of $\log \tau_{\text{lag}}$ and $\log \mathscr{F}$ are also important in the third factor. It means that there is some connection of τ_{lag} with the emitted energies of GRBs, and thus the luminosity indicator based on the spectral lag seems to be indirectly supported by the structure of third factor ([10, 13, 14]). If the Amati-relation ([1]) stands then there should be a linear connection between $\log E_{\text{pk;intrinsic}}$ and $\log E_{iso}$. The Amati-relation was indeed predicted by the strong correlation between $\log \mathscr{F}$ and $\log E_{\text{pk}}$ ([8]). The correlation between $\log \mathscr{F}$ and $\log E_{\text{pk}}$ does not mean that there is a linear connection *only* between $\log E_{iso}$ and $\log E_{\text{pk;intrinsic}}$. In fact, [3] arrived also to the conclusion that

$$\log E_{iso} = a_1 \log E_{\text{pk;intrinsic}} + b_1 \log \tau_{intrinsic} + c_1 \qquad (2)$$

should hold with some suitable a_1, b_1, c_1 constants ($\tau_{intrinsic} = \tau/(1+z)$). Note that $\mathscr{T}_{50}$ and τ strongly correlates with each other, i.e., in this equation either τ_{intr} or $\mathscr{T}_{50;intrinsic}$ can be used. Recently, the validity of the Amati-relation has been a matter of intensive discussion ([14, 9, 5, 4]). The factor loadings show that $\log \mathscr{F}$ is explained basically by the first three factors. Since Fact. 1 is mainly given by $\log \tau$, Fact. 2 by $\log F_{1s}$ and Fact. 3 by $\log E_{\text{pk}}$, all this suggests that a relation of the form

$$\log E_{iso} = a_2 \log E_{\text{pk;intrinsic}} + b_2 \log \tau_{intr} + c_2 \log L_{iso} + d \qquad (3)$$

should exist, with some suitable a_2, b_2, c_2, d constants. Note that a similar relation was proposed also by [5]. It follows from the first three factors that the relationship of $\log \mathscr{F}$ and $\log E_{\text{pk}}$ is less important than that of the variables dominating Fact. 1 and 2, because $\log \mathscr{F}$ and $\log E_{\text{pk}}$ together are mainly determined by Fact. 3, and thus for their relation one cannot omit the variables that are dominating Fact. 1 and Fact. 2, respectively. This fact disfavors a simple linear relationship *only* between $\log E_{\text{pk;intrinsic}}$ and $\log E_{iso}$.

The **fourth factor** is dominated by V, $\mathscr{S}_{\mathscr{F}}$ and $\mathscr{R}E_{\text{pk}}$. However, according to [11] and [12], the variability should be coupled to the luminosities of GRBs, and hence to the fluence and peak flux. No such connection is supported by the fourth factor. Hence, some queries emerge here for the redshift estimations derived from the variability.

The **fifth factor** is dominated by T_{90} and $\mathscr{T}_{50}$. This shows that T_{90} and $\mathscr{T}_{50}$ are not completely equivalent, though $\mathscr{T}_{50}$ better characterizes a burst.

CONCLUSIONS

The results of the paper may be summarized as follows.

- No more than 5 factors should be introduced. This essential lowering of the significant variables is the key result of this paper.
- The structure of factors is similar to the PCs of [3]. The number of important quantities is more accurately defined here.
- The first factor is given mainly by the temporal variables, and the quantities τ and $\mathcal{T}_{50}$ are preferred.
- The second factor is related to the strength of the burst.
- The connection of E_{pk} in the third factor with other quantities, and the structure of the first three factors cast considerable doubts about the Amati-relation in its original form. For the luminosity indicators based on the spectral lag some support emerges from the third factor.
- The variability in fourth factor does not support its connection to the intrinsic luminosities, and the pseudo-redshift estimations based on the variability.
- The fifth factor shows that T_{90} and $\mathcal{T}_{50}$ are not completely equivalent.

Because all these conclusions are obtained from the measured data alone, all models of long GRBs must respect these expectations.

ACKNOWLEDGMENTS

REFERENCES

1. Amati, L., et al. 2002, A&A, **390**. p. 81
2. Bagoly, Z., Mészáros, A., Horváth, I., Balázs, L.G. & Mészáros, P. 1998, ApJ, **498**. p. 342
3. Borgonovo, L. & Björnsson, C.-I. 2006, ApJ, **652**. p. 1423
4. Butler, N.R., Kocevski, D., Bloom, J.S. & Curtis, J.L. 2007, astro-ph/0706.1275
5. Firmani, C., Ghisellini, G., Avila-Reese, V. & Ghirlanda, G. 2006, MNRAS, **370**. p. 185
6. Hakkila, J., et al. 2003, ApJ, **582**. p. 320
7. Kendall, M.G. & Stuart, A. 1973, The Advanced Theory of Statistics, *Charles Griffin & Co. Ltd., London & High Wycombe*
8. Lloyd, N. M., Petrosian, V. & Mallozzi, R. S. 2000, ApJ, **534**. p. 227
9. Nakar, E. & Piran, T. 2005, MNRAS, **36**. p. L73
10. Norris, J.P. 2002, ApJ, **579**. p. 386
11. Ramirez-Ruiz, E. & Fenimore, E.E. 2000, ApJ, **539**. p. 712
12. Reichart, D. E., et al. 2001, ApJ, **552**. p. 57
13. Ryde, F. 2005, A&A, **429**. p. 869
14. Ryde, F. 2005, ApJL, **625**. p. 95